ESTIMATION OF GREAT LAKES WATER LEVEL STATISTICS: CONDITIONING VIA "THE BOOTSTRAP"

*Lynn R. Herche and Holly C. Hartmann**

ABSTRACT: Reliable lake level frequency distributions are a critical component of any comprehensive strategy for coping with Great Lakes water level fluctuations. However, statistical techniques commonly used on riverine systems are inappropriate for large lake systems, due to the levels' long-term persistence and dependence on the prevailing climatic regime. To illustrate an alternative methodology, we present a series of resampling analyses modeled after well-known bootstrap techniques, applied to 130 years of monthly Lake Erie water level records. The analyses show that lake level exceedance probabilities should be conditioned on 1) length of planning horizon, 2) starting month of planning horizon, 3) initial lake level, and 4) climatic regime. Our methodology can be extended to additionally consider storm and wind effects on levels, to incorporate levels data available for discontinuous periods prior to 1860, and to develop other types of lake level statistics useful to decision makers, such as duration and time-to-exceedance probabilities.

Introduction

The Great Lakes are one of the most intensively used freshwater systems in the world, serving navigation, hydropower, irrigation, water supply, and recreation interests, while providing important fish and wildlife habitats. Due to their large surface areas and relatively small outflow capacities, the lakes fluctuate through a very small range of levels compared to smaller lake or riverine systems (historically, about 6 ft [1.83 m] from record lows to record highs). In addition, changes in lake levels are typically gradual from year to year. Thus, uses of the lakes have generally evolved to accommodate a relatively narrow range of lake level conditions. While Great Lakes uses are adapted to seasonal fluctuations, extreme high and low water events and rapid level changes pose major management challenges. Reliable lake level probability distributions are a critical component of any comprehensive strategy for coping with Great Lakes water level fluctuations. Such distributions permit consideration of the risks associated with investment decisions involving lake resources.

The wide variety of Great Lakes uses implies a need for a wide variety of statistics describing potential lake level fluctuations. Shoreline development interests are most concerned with instantaneous peak levels, since any inundation can cause significant damage. Water intake design and operation is concerned not just with extreme levels, but also with the expected duration of specific extremes, especially at severe low levels. Because peak hydropower demands typically don't occur concurrently with peak water levels, statistics about expected durations at even moderate levels are of interest as well. In addition, various Great Lakes uses have different planning horizons. Major public works may have design lifetimes of 50 years or longer. On the other hand, new marina operations may have a critical planning horizon of 2 years or less, since the business may fail if extreme conditions occur before financial reserves have been accumulated.

**Great Lakes Environmental Research Laboratory, National Oceanic and Atmospheric Administration, Ann Arbor, MI*

Although Great Lakes water level measurements comprise among the longest North American geophysical instrumental records, it is inappropriate to use them directly to create probability distributions based on techniques developed for riverine systems. Great Lakes levels are highly serially-correlated due to the tremendous heat and moisture storage capacities of the lakes and their basins, respectively, and the restricted lake outlets. In addition, historic lake level records reflect secular changes in climate, watershed hydrologic response, and connecting channel hydraulics. Such conditions violate assumptions of independent, identically distributed events essential to traditional statistical hydrologic analyses. Improved methodologies are needed for producing lake level frequency distributions that consider periodic climatic shifts, the long lag-response of the lakes to meteorologic variability, and current hydrologic and hydraulic conditions.

Attempts to address constraints in applying traditional statistical techniques directly to water level records of other large lakes generally focus on analyses of net lake inflows or their components (Guganesharajah and Shaw 1984, James et al. 1984, Adams et al. 1985, Bowles and James 1985, Wiche et al. 1986, Privalsky 1981, 1988). However, direct application of these approaches to the Great Lakes is complicated, because, except for Lake Superior, the largest water supply to each Great Lake is outflow from its upstream lake; being a direct reflection of lake level conditions, those outflows are subject to the same long-term persistence as water levels. Because lake level changes reflect long-term meteorologic variation, lake level frequency distributions would be, ideally, derived from distributions of meteorologic variables, combined with modeling of hydrologic and hydraulic processes. However, such an approach requires a long-term effort, since no applicable meteorologic probability distributions exist for the Great Lakes region at present and existing models are not fully developed for such use.

Many Great Lakes investment decisions cannot wait for development of "ideal" lake level probability distributions. The U.S. federal flood insurance program requires estimation of annual flood probabilities (FEMA 1987) and affects a broad range of shoreline development. The recent spate of reports (Bishop 1987, Hartmann 1987, USACE 1988, SWRPC 1989) suggesting the potential for future lake level variations, developed in light of the extreme levels of the 1980s, reflects demands for statistics that can be used now and that are more reliable than those available previously.

This paper represents an attempt to develop techniques for improved Great Lakes level probability distributions, that can be applied in a timely manner to better guide Great Lakes investment decisions over the near future. Our focus is on using solely lake level records; as Quinn (1990) points out, there is a general view that existing analyses haven't sufficiently exploited the information contained in historic levels records. We present a series of resampling analyses modeled after well-known bootstrap techniques, applied to 130 years of monthly Lake Erie water level records. The probability distributions presented herein are not intended to be used directly for design, since there have been no adjustments to historic levels for diversions, connecting channel hydraulics, or lake outflow regulation. Rather, our analyses are presented as an illustration of methodology, and to make clear the problematic implications of ignoring the physical realities of large lake behavior.

Methodology

Our resampling analyses use a single continuous 130-year (1860-1989) record of monthly mean water levels on Lake Erie, recorded at Cleveland, Ohio. This record (Figure 1) was thought to be appropriate for three reasons. First, 130 years is quite long for such a data record and its continuity permitted a variety of methods to be used. Second, Cleveland is located on the southern shore of Lake Erie near the midpoint of the long east-west axis of the lake and, hence, is much less subject to the periodic water

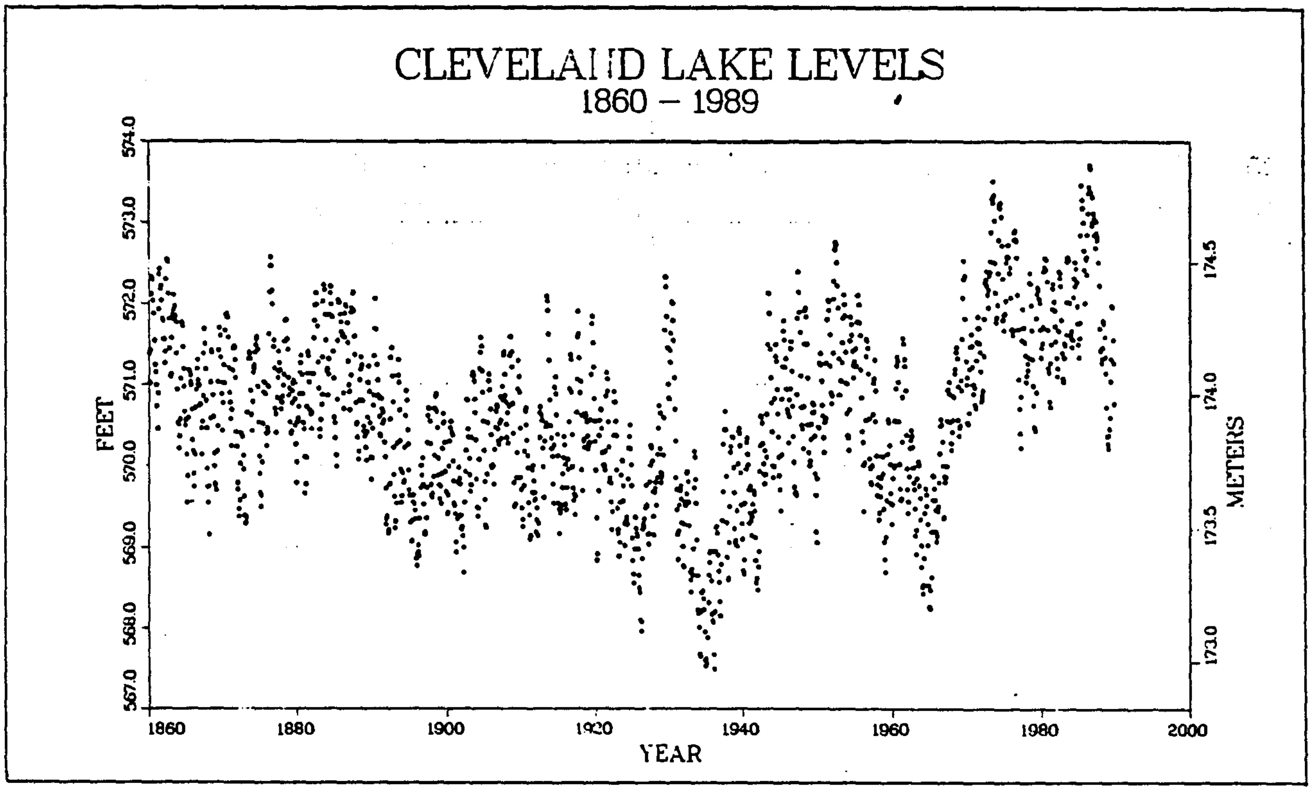

Figure 1. Historic monthly mean Lake Erie levels at Cleveland, Ohio.

level extremes or seiches induced by relatively short-term storm and wind events. Third, of the five major Great Lakes, Lake Erie levels were thought to have been least subject to human effects such as regulation and dredging over the period of record. Thus the data for analysis is this set of monthly mean lake levels, denoted L_i, i = 1-1558.

The monthly lake level time series exhibits two well known features. The first is a very strong positive correlation existing between levels near in time. Lake levels are clearly not independent events. Second, a major seasonal effect is present with levels generally being highest in May through July and lowest in November through January. This reflects the seasonality of hydrologic processes (basin runoff and lake evaporation) as they respond to seasonal meteorologic conditions. Additionally, a third feature, not immediately apparent from the levels record alone, is the strong positive relation between lake level and rate of discharge from the lake. While an average water supply occurring at a low lake level causes the lake to rise, that same supply occurring at a high level results in a drop in levels.

These features suggest certain approaches to be applied in the analyses. The first characteristic, that the levels are strongly positively correlated, suggests that we analyze not the levels themselves, but their differences. Lake level differences from month to month reflect more directly the net effects of short-term meteorologic variability (via overlake precipitation, basin runoff, and lake evaporation). We define a forward difference at time i between levels L_{i+1} and L_i as $\Delta_i = L_{i+1} - L_i$. This difference series has very small (.005 in absolute value) and non-significant ($p > 0.1$) auto-correlations. Woodbury and Padmanabhan (1989) used a differencing approach to circumvent the long-term persistence of levels at Devils Lake, North Dakota; they analyzed time series of incremental storage differences based on annual lake levels using ARMA models, and annual maximum deviations from average annual levels using extreme value theory.

The second and third features suggest a structure of grouping the differences by month and level, as illustrated in Figure 2. The range of levels is divided into 10 intervals using nine division points 0.5 ft (.1524 m) apart. The Δ_i's are grouped by the interval and month to which the corresponding L_i belong. The Δ_i's occurring at or near the same lake level would have been subject to similar effects of attenuation or emphasis due to rate of discharge from the lake. Similarly, Δ_i's occurring in the same portion of the annual meteorologic cycle would be, in general, subject to similar effects of hydrometeorologic processes. The strong seasonal variation in levels in evident in Figure 2. Guganesharajah and Shaw (1984)

Observations by Level and Month

Mon: / Level(ft)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
	1	3	7	8	10	13	11	8	2	1	2	2
572.5	2	3	2	11	18	19	15	12	10	3	1	1
572.0	9	5	9	20	18	25	27	23	14	12	9	6
571.5	12	15	13	12	26	26	30	28	24	20	12	12
571.0	17	14	23	29	23	18	18	23	31	24	21	25
570.5	24	22	20	20	16	15	13	20	23	28	27	18
570.0	28	26	24	19	13	9	11	8	13	24	29	32
569.5	24	27	19	7	4	3	3	5	9	9	16	19
569.0	6	11	9	2	1	1	1	2	3	7	7	7
568.5	7	4	4	2	1	1	1	1	1	2	6	8

Figure 2 . Grouping of historic levels differences by month and level, for the period 1860-1989.

used an interval approach in estimating probability distributions of annual minimum 10-day water levels for Lake Chad, conditioned on initial lake levels. They designated 12 discrete intervals of levels, each covering 1.64 ft (0.5 m). Starting with the mid-point of each interval, in turn, they used an autoregressive model of lake inflows to generate multiple sequences of lake levels, and ultimately determined the proportion of occurrences of levels falling within each of the 12 intervals, for the specified starting condition.

The method chosen here to analyze the levels, via their differences, is closely related to the "bootstrap" method (Efron 1982). In perhaps the most familiar form, one is given a sample $X = (x_1, x_2, \ldots x_n)$ of observations from a typically unknown distribution, $F(x)$, and a statistic, $g(X)$. The objective is to estimate some distributional characteristic of $g(X)$ such as its standard error, $S(g)$. The method is summarized as:

1. Fit the non-parametric maximum likelihood estimate of F: $\hat{F}$, which assigns mass $1/n$ at x_i, $i = 1,\dots,n$.
2. Draw, with replacement, a "bootstrap sample" from $\hat{F}$: $X^* = (x_1^*,\dots,x_n^*)$ where x_i^* are independently and identically distributed according to $\hat{F}$.
3. Calculate $\hat{g}^* = g(X^*)$; i.e., compute $g(X^*)$ from the bootstrap sample.
4. Independently repeat steps 2 and 3 a large number, B, of times, getting bootstrap replications $\hat{g}^{*1},\dots,\hat{g}^{*B}$.
5. Finally, calculate the desired characteristic of g, such as its estimated standard error,

$$\left\{ \sum_{b=1}^{B} [\, \hat{g}^{*b} - \hat{g}^{*\bullet} \,]^2 / (B-1) \right\}^{1/2}$$

where $\hat{g}^{*\bullet} = \left(\sum_{b=1}^{B} \hat{g}^{*b} \right) / B$.

In contrast with this illustration using $S(g)$, the most useful applications of bootstrap techniques are typically found in situations for which analytic results are essentially intractable or unknown. Heuristic application of the bootstrap method to the lake levels data proceeds with these steps:

1. Pick a starting level, l_1, and starting month, m_1. All subsequent operations will thus be conditional on the selected starting level and month.

2.a) Randomly draw a δ_1 from the set of Δ's corresponding to l_1 and m_1.

2.b) Compute a new level, $l_2 = l_1+\delta_1$ and step to the next month, $m_2 = m_1(\mathrm{mod}12) + 1$.

3. Repeat steps 2a) and 2b) for $l_{i+1} = l_i+\delta_i$ and $m_{i+1} = m_i(\mathrm{mod}12) + 1$, as i increments from 2 to the number of months in the planning horizon of interest. (For example, next draw a δ_2 from the l_2,m_2 set of Δ's and compute $l_3 = l_2+\delta_2$ and $m_3 = m_2(\mathrm{mod}12) + 1$. Continue this for the time of interest, drawing 24 δ's for a 2-year planning horizon, say, and finish with a set of 25 l_i's.)
4. Collect the statistics of interest for the set of l_i's. (For example, the maximum of the 25 l_i's.)
5. Repeat steps 2 through 4 a large number, B, of times. (In our example, this results in B maxima, one from each iteration of the steps 2 - 4.)

The resultant collection of B resampling statistics reflects the distribution of the desired statistic computed on the B samples. Figure 3 diagrams the construction of one of the B resamplings and determination of the statistic of interest (here, the maximum level achieved during the chosen planning horizon, conditioned on a starting level and month). One may also take the minimum of the 25 levels to assess

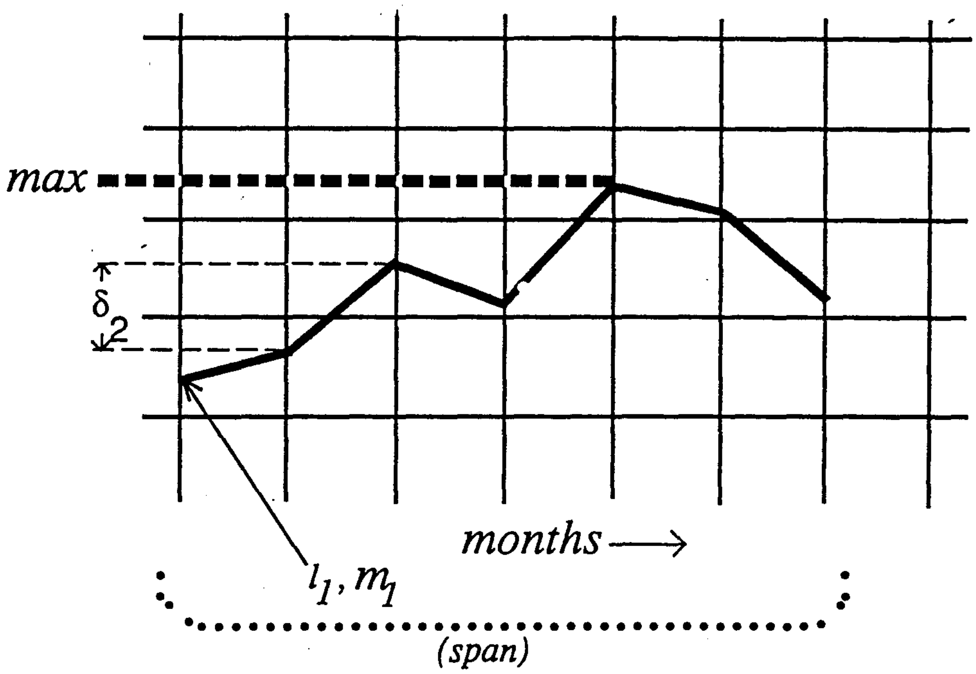

Figure 3. Schematic of resampling of lake level differences from levels intervals on the basis of month and lake level. Solid line shows sequence of lake levels resulting from one sampling replication (B=1) over a 6-month period. Regardless of the length of the sampling period, all sampling is thus conditioned on the initial lake level and month.

the distribution of minima under the same conditioning. Note that the conditioning which takes place in the choice of starting level, l_1, and month, m_1, is constant over the B samples. In the analyses which follow, we use B=30,000; a preliminary analysis indicated stable results with B > 15,000.

As an example, Figure 4 shows the result of conditioning on an arbitrary starting level of 570.75 ft (173.96 m) in January, with a sampling span, or planning horizon, of one year. Point A indicates that approximately 300 of the 30,000 samples achieved a maximum level of 572.75 ft (174.57 m) or above during the 1-year span. Thus, Figure 4 shows a 1% probability that levels will exceed 572.75 ft (174.57 m) over a 1-year planning horizon, when the monthly mean lake level at the start of the period is at 570.75 ft (173.96 m). Point B indicates approximately 30 of the samples achieved a minimum level of 568.9 ft (173.39 m) or below during the period. Thus, for the specified conditional starting level, there is a 99.9% probability that levels over the next year will exceed 568.9 ft (173.39 m); alternatively, there is a 0.1% probability that levels over the period will not exceed 568.9 ft (173.39 m). Note that this example is for a 1-year planning horizon. For other spans, the curves represent the probability that indicated levels will be exceeded at some unspecified time during the planning horizon; they do not represent annual exceedance probabilities.

Conditioning in this type of analysis corresponds in kind to using exogenous variables in regression models. Two conditioning factors have been included in the sampling structure, namely the seasonal and levels effects. The natural sequencing of seasonal effects and the dependency of a response (Δ) on

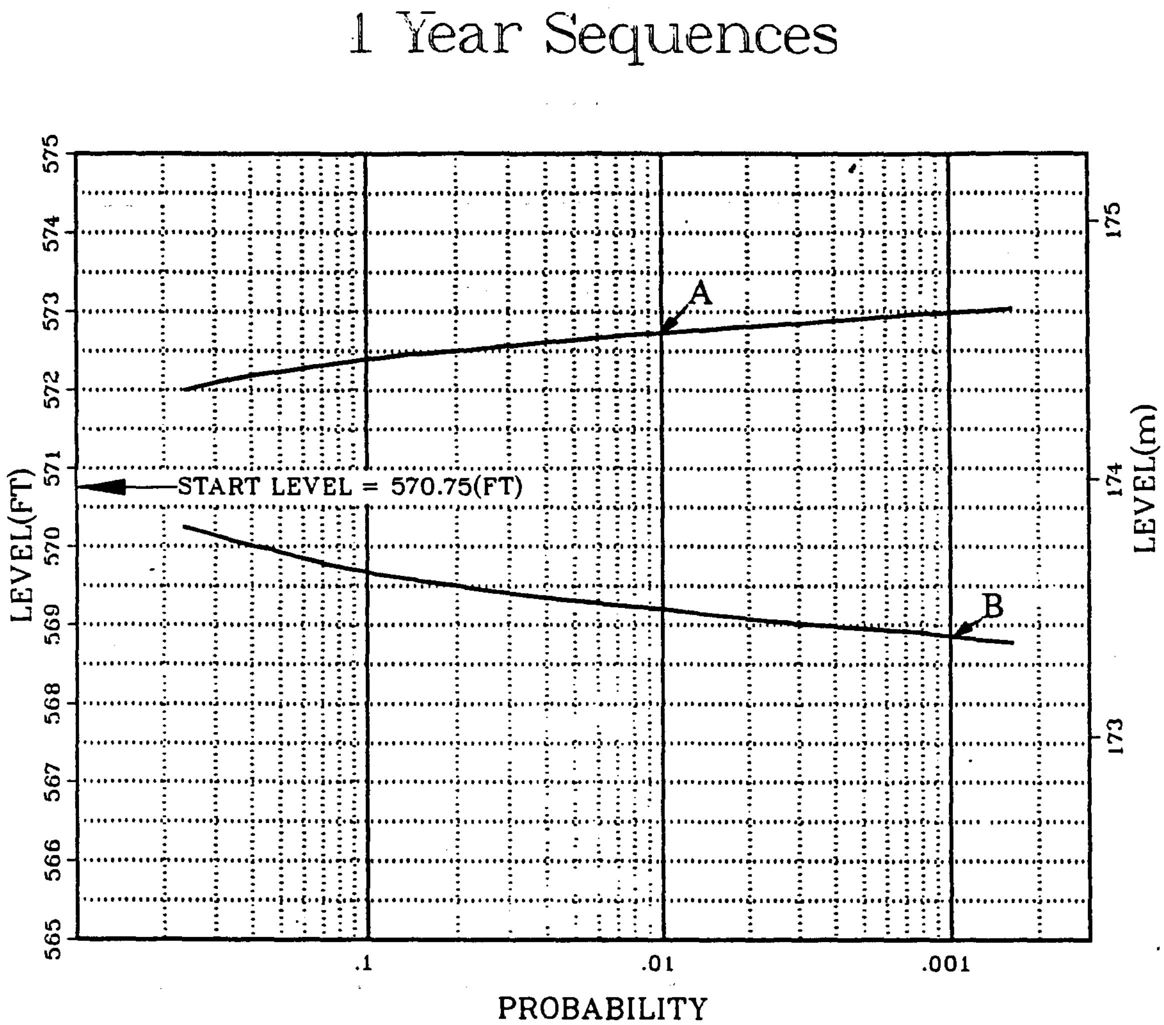

Figure 4. Example probability distributions for monthly mean Lake Erie levels at Cleveland, Ohio, over a 1-year planning horizon, conditioned on an initial January lake level of 570.75 ft. (173.96 m). A=lake level with 1% exceedance probability; B=level with 0.1% non-exceedance probability.

an associated level essentially dictate the structure of choosing elements (δ's) of a sample sequence. The analyses which follow illustrate the effects of applying various kinds of conditioning, as well as some interesting interactions.

Results

Even without any consideration of conditioning, differences in planning horizons can cause a strong variation in maxima or minima. Figure 5 shows the variation in four different periods, starting in January at 572.75 ft (174.57 m) and increasing in 6-year steps from 1 to 7, 13, and 19 years. Two major features are evident here. First, the effects of planning horizon differences weaken as the period increases, and may become negligible for periods of more than 20 years, depending on starting level. The maxima are essentially unchanged beyond the 7-year planning horizon for this rather high starting level. The changes in minima probabilities are very large between the 1- and 7-year planning horizons, but decrease rapidly as the planning horizon extends beyond 13 years. In the example of Figure 5, the

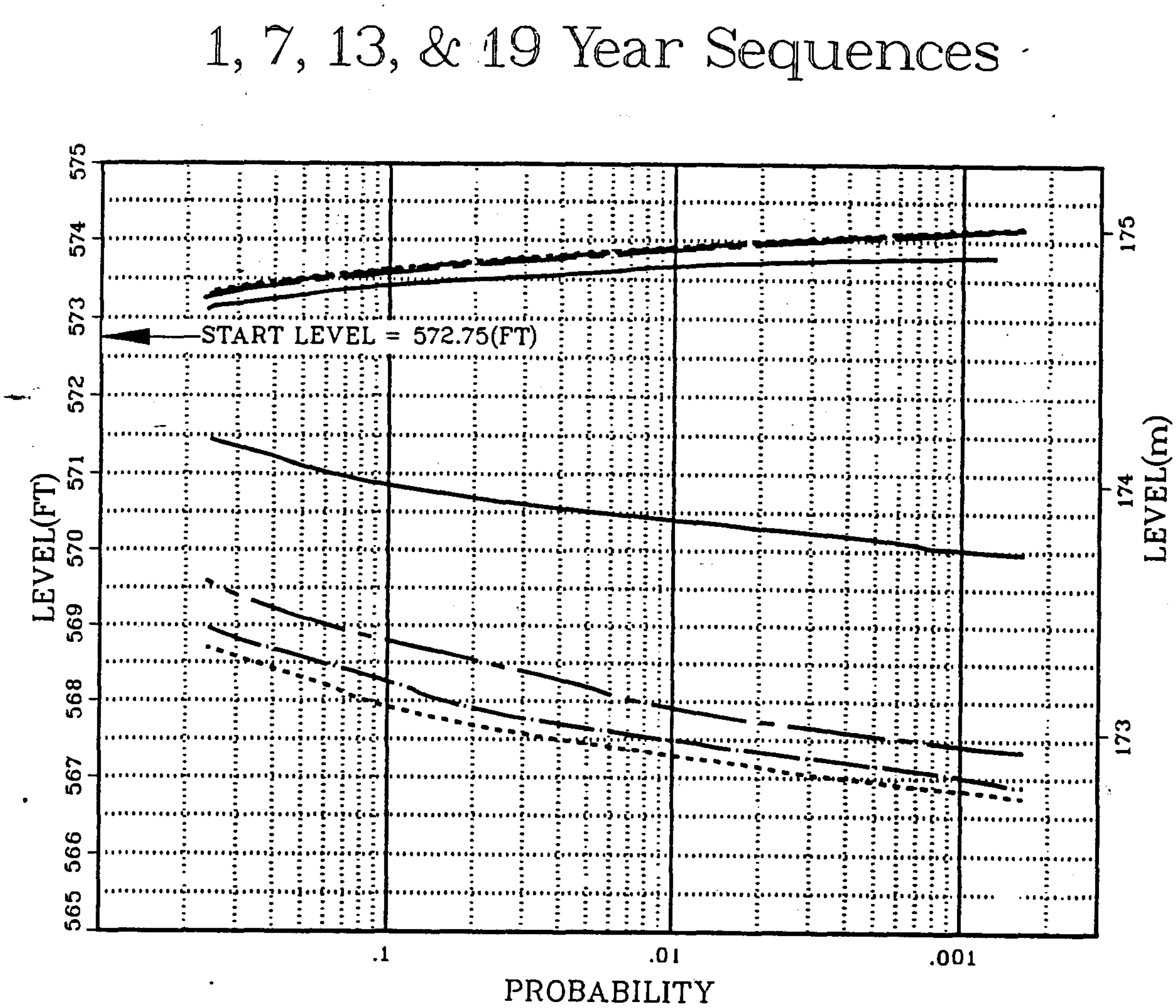

Figure 5. Example probability distributions for monthly mean Lake Erie levels at Cleveland, Ohio, over 1-, 7-, 13-, and 19-year planning horizons, conditioned on an intial January lake level of 572.5 ft. (174.49m). The planning horizons are distinguished by the solid, long-dashed, short-dashed, and dotted lines, respectively.

1% non-exceedance probability lake level for the 1-year planning horizon is about 2.5 ft (0.76 m) higher than for the 7-year planning horizon, and over 3 ft (0.91 m) higher than for the 19-year period. Second, the marked asymmetry between minima and maxima probability curves is largely due to the high January starting level. This occurs because at high lake levels, even above-average water supplies can result in lake level drops due to large lake outflows. Thus, few resamplings of lake level differences within the high lake level intervals of Figure 2 yielded lake level increases and those that did occur were relatively minor. However, many resamplings within the high levels intervals yielded lake level drops. Unless conditioned on the length of the planning horizon, the likelihood of extreme conditions over short planning horizons will be overestimated; if lake levels are currently extreme, the probability of extremes of the opposite sort will be grossly overestimated.

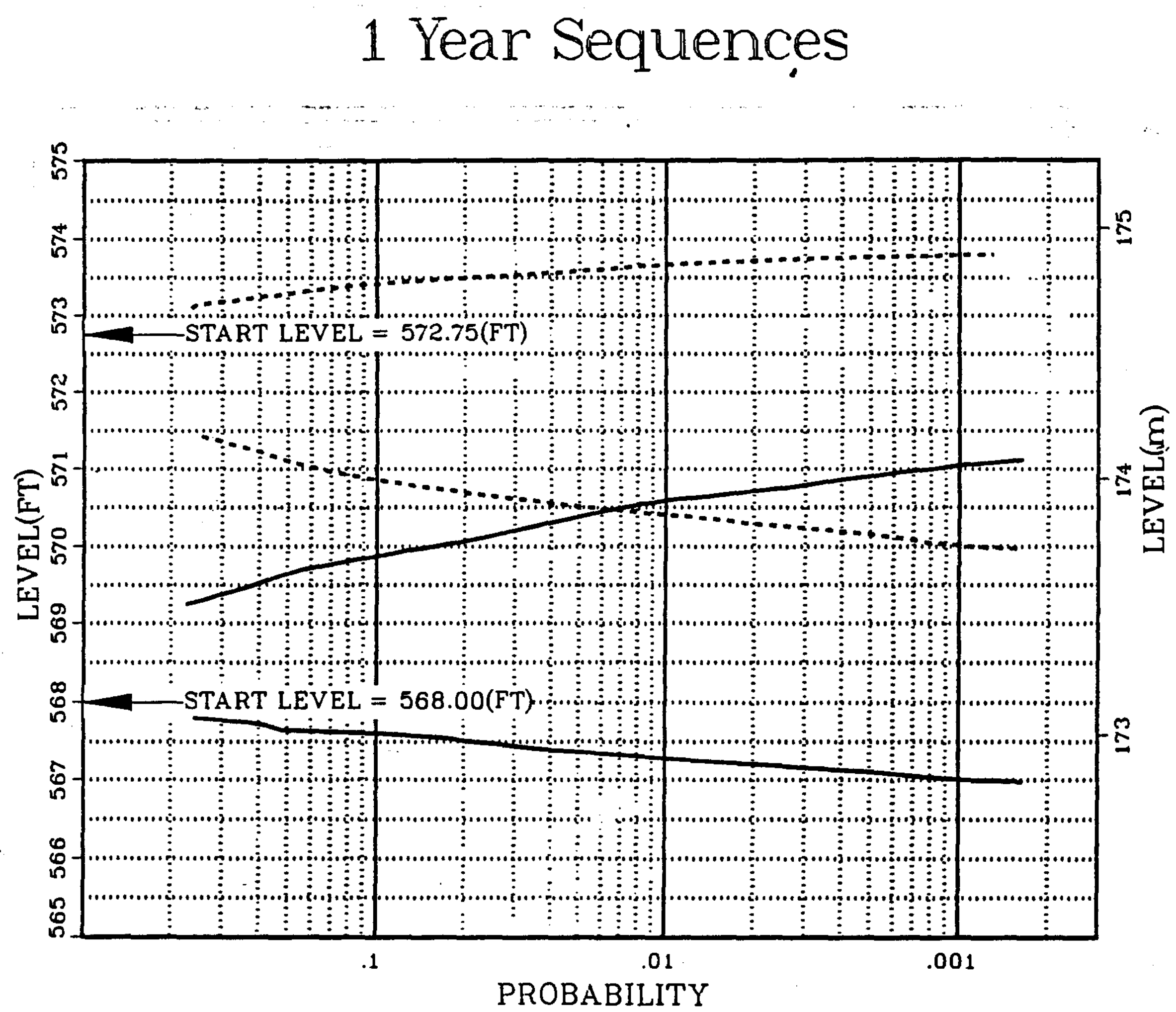

Figure 6. Example probability distributions for monthly mean Lake Erie levels at Cleveland, Ohio, over a 1-year planning horizon, conditioned on two different initial January lake levels. Dotted lines are conditioned on initial levels of 572.75 ft (174.57 m), solid lines on initial levels of 568.0 ft (173.11 m).

Compare now, in Figures 6, 7, and 8 for 1-, 10-, and 20-year planning horizons, respectively, the minima and maxima probabilities with two rather extreme starting levels of 568.0 ft (173.11 m) and 572.75 ft (174.57 m). Responses are very sensitive to starting levels for short planning horizons. For the 1-year planning horizon of Figure 6, the 1% exceedance probabilities differ by over 3 ft (0.91 m), as do the 1% non-exceedance probabilities. The probability distributions' sensitivity to starting levels diminishes as the planning horizon increases, but is still notable even for 20-year periods. In Figure 8, the 1% non-exceedance probabilities differ by about 0.5 ft (0.15 m), certainly significant for some Great Lakes uses.

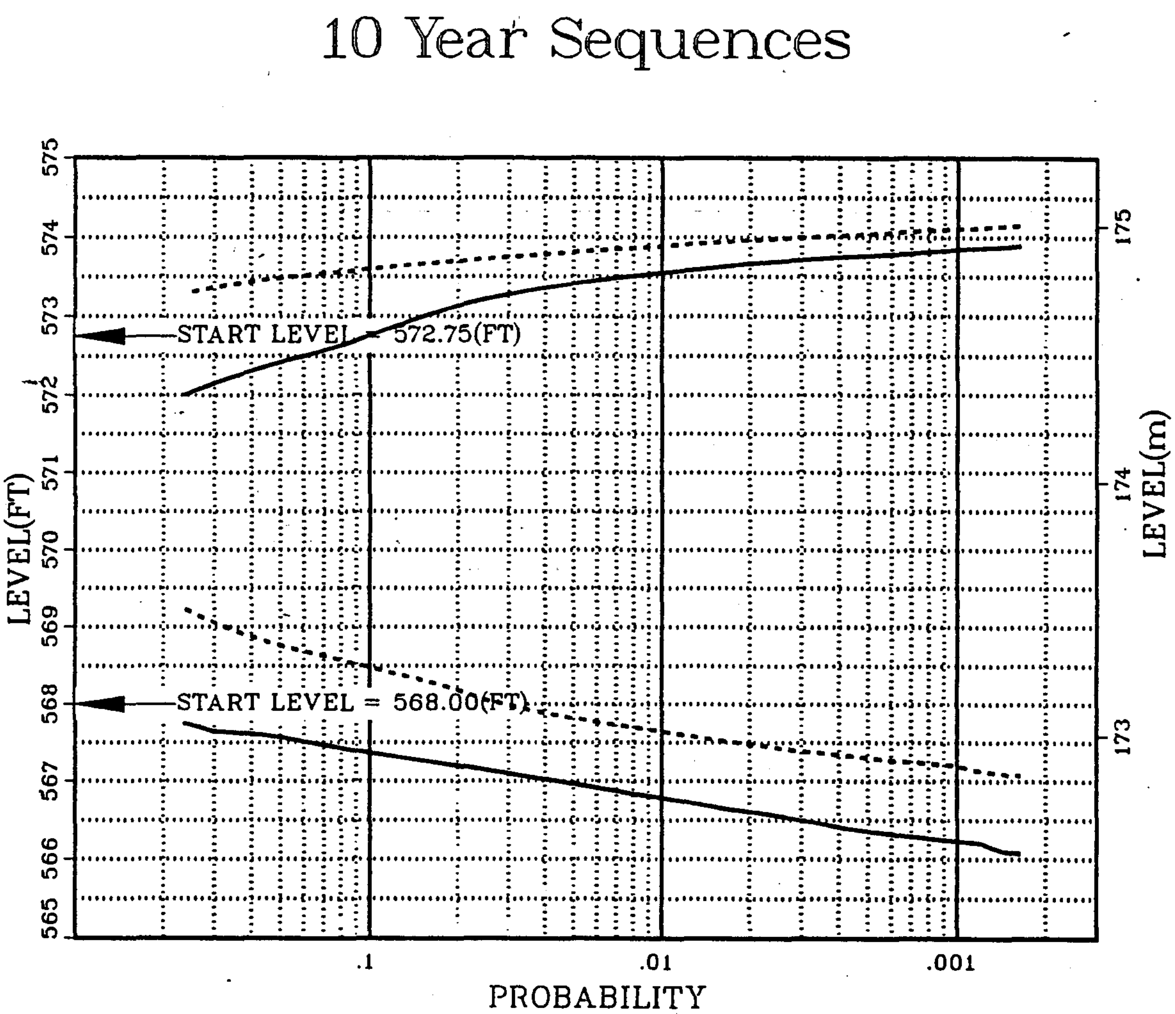

Figure 7. Example probability distributions for monthly mean Lake Erie levels at Cleveland, Ohio, over a 10-year planning horizon, conditioned on two different initial January lake levels. Dotted lines are conditioned on initial levels of 572.75 ft (174.57 m), solid lines on initial levels of 568.0 ft (173.11 m).

Probability distributions that are conditional on the month in which they begin show an analogous pattern of influence. In Figure 9, the higher maximum and minimum curves started at 570.75 ft (173.96 m) in January when levels are generally lower. The lower maximum and minimum probability curves started at the same lake level but in June when levels are generally higher. As with the preceding conditioning factors, the sensitivity to the starting time weakens as the span increases. Over 12-year periods, Figure 9 shows relatively minor differences between 1% probability levels, although differences are somewhat larger for lake levels of greater probability.

The implications of Figures 5-9 are clear. Development of a single general-purpose probability distribution for monthly mean lake levels is absolutely inappropriate. Use of non-conditional probabilities for short planning horizons of about 20 years or less will result in over-investment in risk reduction measures. Probability distributions must consider: 1) the length of the planning horizon,

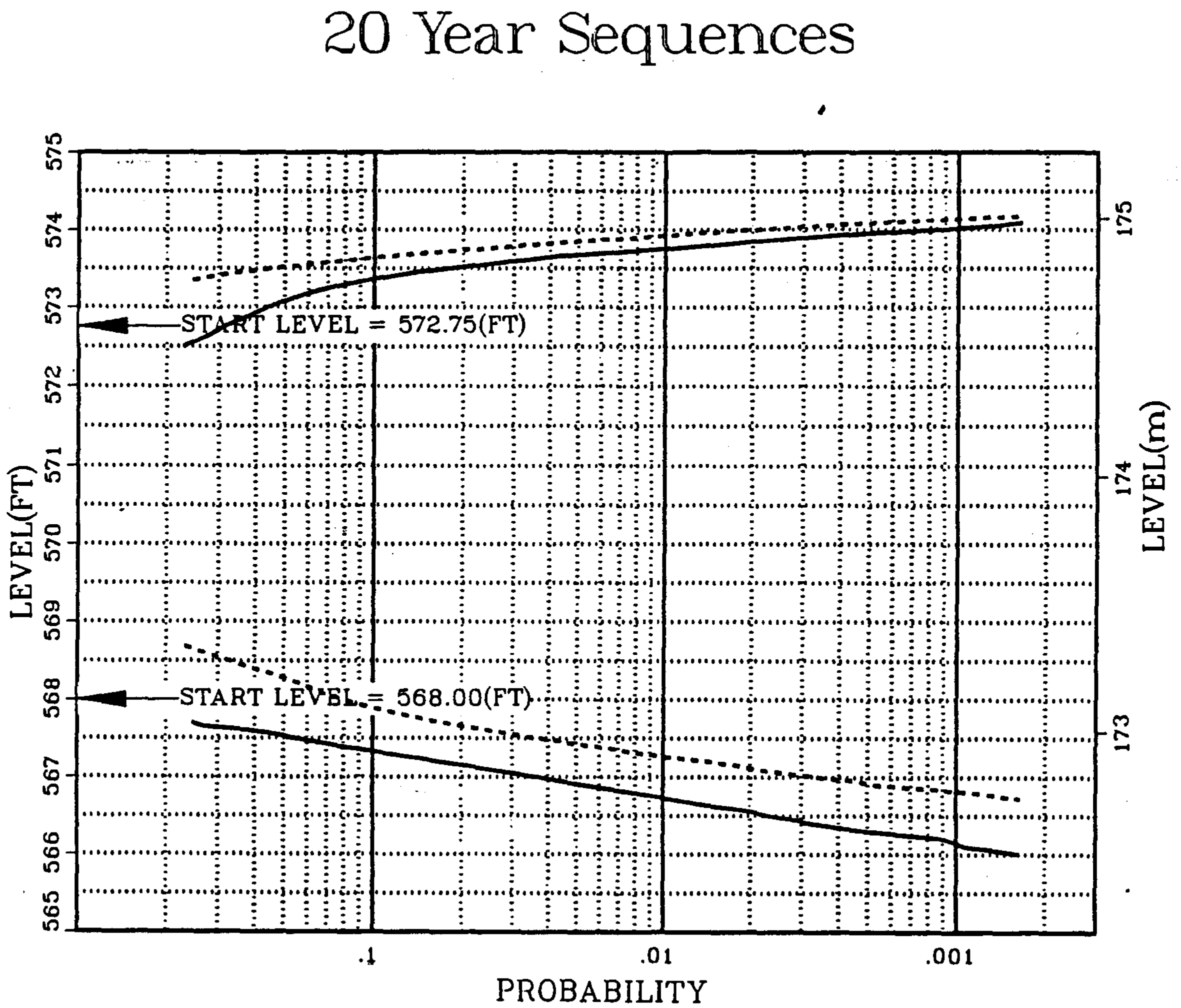

Figure 8. Example probability distributions for monthly mean Lake Erie levels at Cleveland, Ohio, over a 20-year planning horizon, conditioned on two different initial January lake levels. Dotted lines are conditioned on initial levels of 572.75 ft (174.57 m), solid lines on initial levels of 568.0 ft (173.11 m).

2) the lake level at the start of the planning horizon, and 3) the month (or season) of the start of the period. As the planning horizon increases beyond 20 years, conditioning on these factors becomes less important, but may still merit consideration when initial levels are extreme and levels of the opposite extreme are of concern.

Thus far, we have considered conditioning by starting level and starting month, having incorporated by construction the effects of seasonality and level-dependent outflow rates. Another factor that strongly influences lake level extremes is that of climatic regimes, i.e., decadal and longer periods of relatively consistent meteorologic conditions. As illustrated by Quinn (1990), the sequence of annual water supplies, which directly reflects persistent climatic conditions, is a critical determinant of extreme Great Lakes water levels. This understanding formed the basis of analyses under the on-going International Joint Commission Great Lakes Water Levels Reference Study (IJC 1989), whereby 12-year

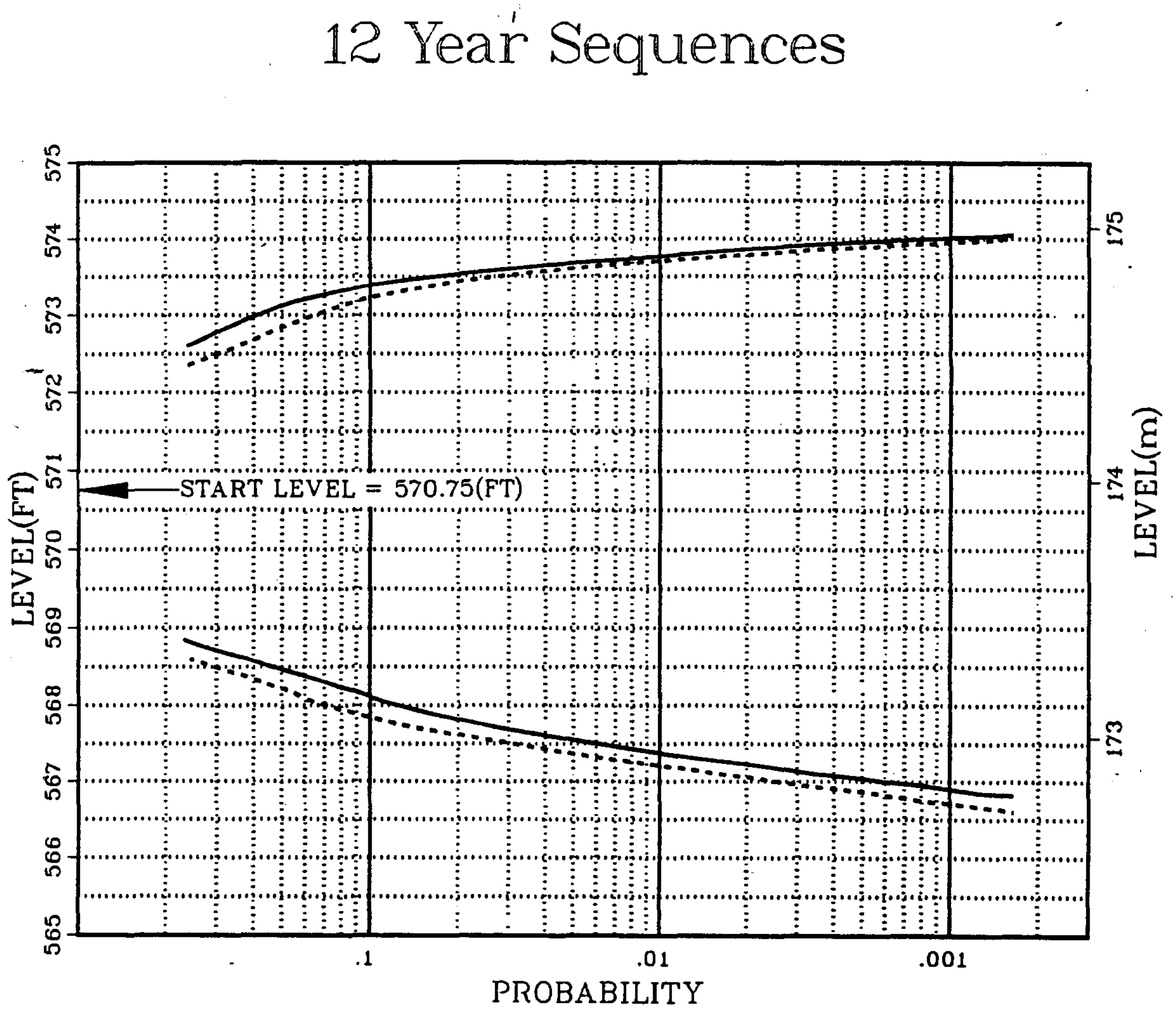

Figure 9. Example probability distributions for monthly mean Lake Erie levels at Cleveland, Ohio, over a 12-year planning horizon, conditioned on January and June initial lake levels of 570.75 ft (173.96 m). Solid lines are conditioned on a January start, dotted lines on a June start.

scenarios of extreme meteorologic conditions based on historic records were used with conceptual models to develop scenarios of lake levels more extreme than reflected by historic levels records; no probabilities of occurrence could be specified for the lake level scenarios, however. The existence of distinct climatic regimes, with relatively rapid shifts between them, is widely acknowledged for the Great Lakes region (Changnon 1987, Quinn 1981, Wiche 1986). For example, over the Lake Michigan basin, cloudiness has increased and temperatures have become cooler with less intra-month variability since the 1960s, while precipitation has been consistently higher since the 1940s (Changnon 1987).

Figure 10 depicts the division of the 130-year Cleveland record into three segments, based on the regimes identified by Quinn (1981; personal communication, GLERL, 1989). The period prior to 1887 and the one following 1941 are classified as high (or wet) regimes while the intervening period is classified as low (or dry). For these analyses, the observations from the two high regimes are pooled

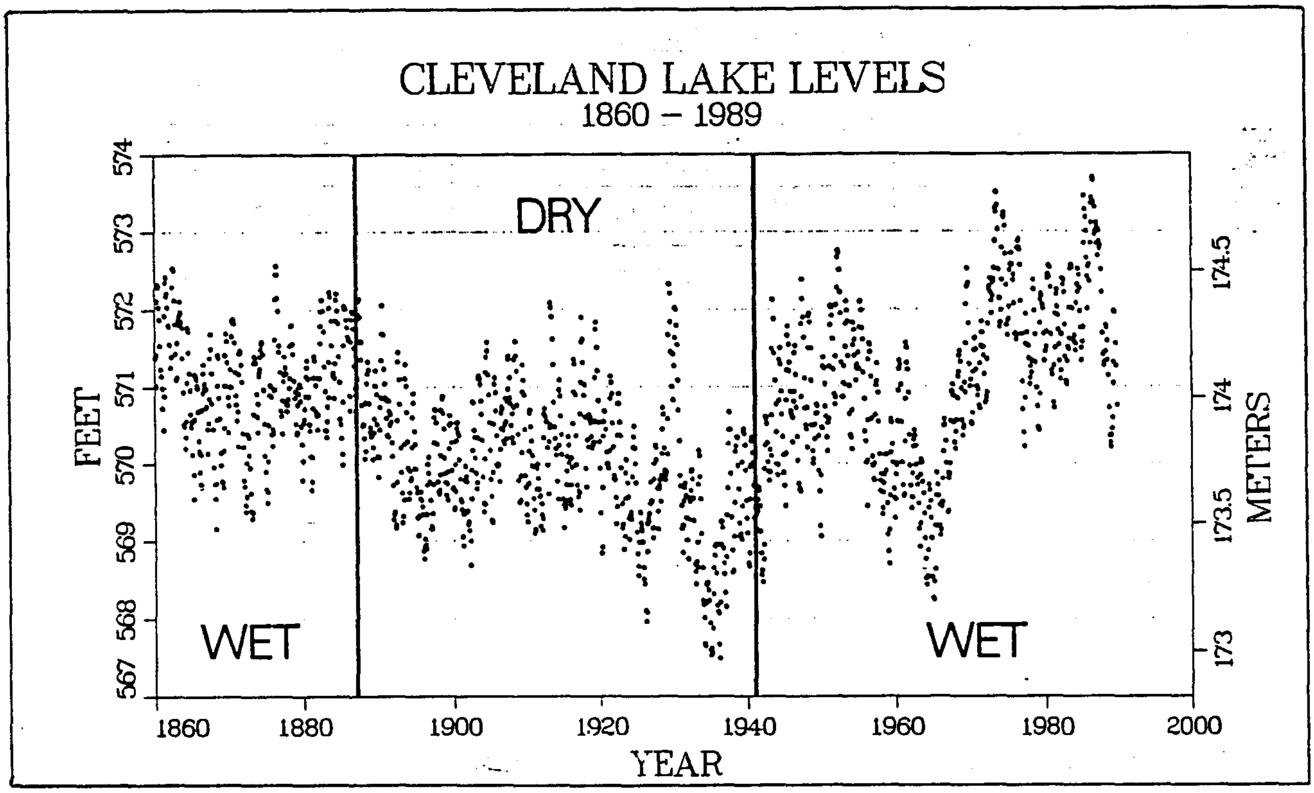

Figure 10. Historic monthly mean Lake Erie levels at Cleveland, Ohio, identified by the prevailing precipitation regime, following Quinn (1981; personal communication, GLERL, 1990).

and used as one regime. Figures 11 and 12 show the numbers of observed lake levels in the two regimes, classified by months and levels. The salient feature of these figures is of course that the observations in the high regime are grouped largely in the higher water levels with complete absence of data at lower levels in late spring and early summer. The low regime data exhibits the complementary pattern with data absent at higher levels, particularly in September through February.

To condition on regime, one performs the same analysis as before, but using data (the Δ's arrayed by level and month) from the regime in question. Figures 13 and 14 show examples for 1- and 12-year planning horizons, respectively, using three data sets: the whole record, the high regime, and the low regime. A central starting level of 570.25 ft (173.80 m) in January was chosen for ease of comparison. The 1-year extremes from this central starting level are all quite close, although the curves from the high regime are consistently slightly above those from the low regime. For the 12-year planning horizon, the predicted extremes are radically different. The 1% non-exceedance probabilities resulting from the wet and dry regimes differ from that obtained using the entire record by about 0.75 ft (0.23 m) and 0.5 ft (0.15 m), respectively. The apparent discontinuity in the maxima probability distribution derived from the dry regime shows the effect of data deficiency at higher lake levels; as shown by Figure 12, for a dry climatic regime, there are few lake level differences at high level intervals from which to resample.

Observations by Level and Month

Level(ft) \ Mon:	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
	1	3	7	8	10	13	11	8	2	1	2	2
572.5												
	2	3	2	9	14	16	14	12	10	3	1	1
572.0												
	9	5	7	18	15	19	21	20	14	12	9	6
571.5												
	11	13	13	7	20	15	19	19	20	19	11	12
571.0												
	14	12	17	22	8	6	3	7	18	17	17	20
570.5												
	15	14	12	5	5	3	4	6	4	13	19	13
570.0												
	15	14	10	5	4	4	4	2	4	4	8	13
569.5												
	4	8	6	2	0	0	0	2	4	4	5	4
569.0												
	1	4	2	0	0	0	0	0	0	2	2	3
568.5												
	3	0	0	0	0	0	0	0	0	0	1	1

High Regime:(1860-1886)+(1942-1989)

Figure 11. Grouping of historic levels differences by month and level, for the combined wet climatic regimes of 1860-1886 and 1942-1989.

Observations by Level and Month

Level(ft) \ Mon:	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
	0	0	0	0	0	0	0	0	0	0	0	0
572.5												
	0	0	0	2	4	3	1	0	0	0	0	0
572.0												
	0	0	2	2	3	6	6	3	0	0	0	0
571.5												
	1	2	0	5	6	11	11	9	4	1	1	0
571.0												
	3	2	6	7	15	12	15	16	13	6	4	5
570.5												
	9	8	8	15	11	12	9	14	19	15	8	5
570.0												
	13	12	14	14	9	5	7	6	9	20	21	19
569.5												
	20	19	13	5	4	3	3	3	5	5	11	15
569.0												
	5	7	7	2	1	1	1	2	3	5	5	4
568.5												
	4	4	4	2	1	1	1	1	1	2	4	6

Low Regime: 1887 - 1941

Figure 12. Grouping of historic levels differences by month and level, for the dry climatic regime of 1887-1941.

The difficulty in using lake level probabilities conditioned on specific climatic regimes is knowing, a priori, what regime will exist over the planning horizon. There is, at present, no ability to predict when climatic regimes will shift, or to immediately identify whether a year that appears to be "unusual" in the context of a prevailing regime is just that (unusual) or is instead the beginning of a new regime. However, as Figures 13 and 14 show, use of the entire historic record to determine level probabilities is not sufficient; Great Lakes levels reflect climatic regimes, and those regimes have different implications for potential lake level behavior. Decision makers should consider lake level probabilities for each regime within their decision analysis framework, by weighted combination of probabilities or some other approach. For extremely long planning horizons (e.g., 100 years for massive infrastructures, 1000 years or beyond for lakeshore nuclear plant siting), several climatic shifts might be expected; thus, use of the entire historic record may be sufficient, or more appropriately, additional conditioning on regime behavior.

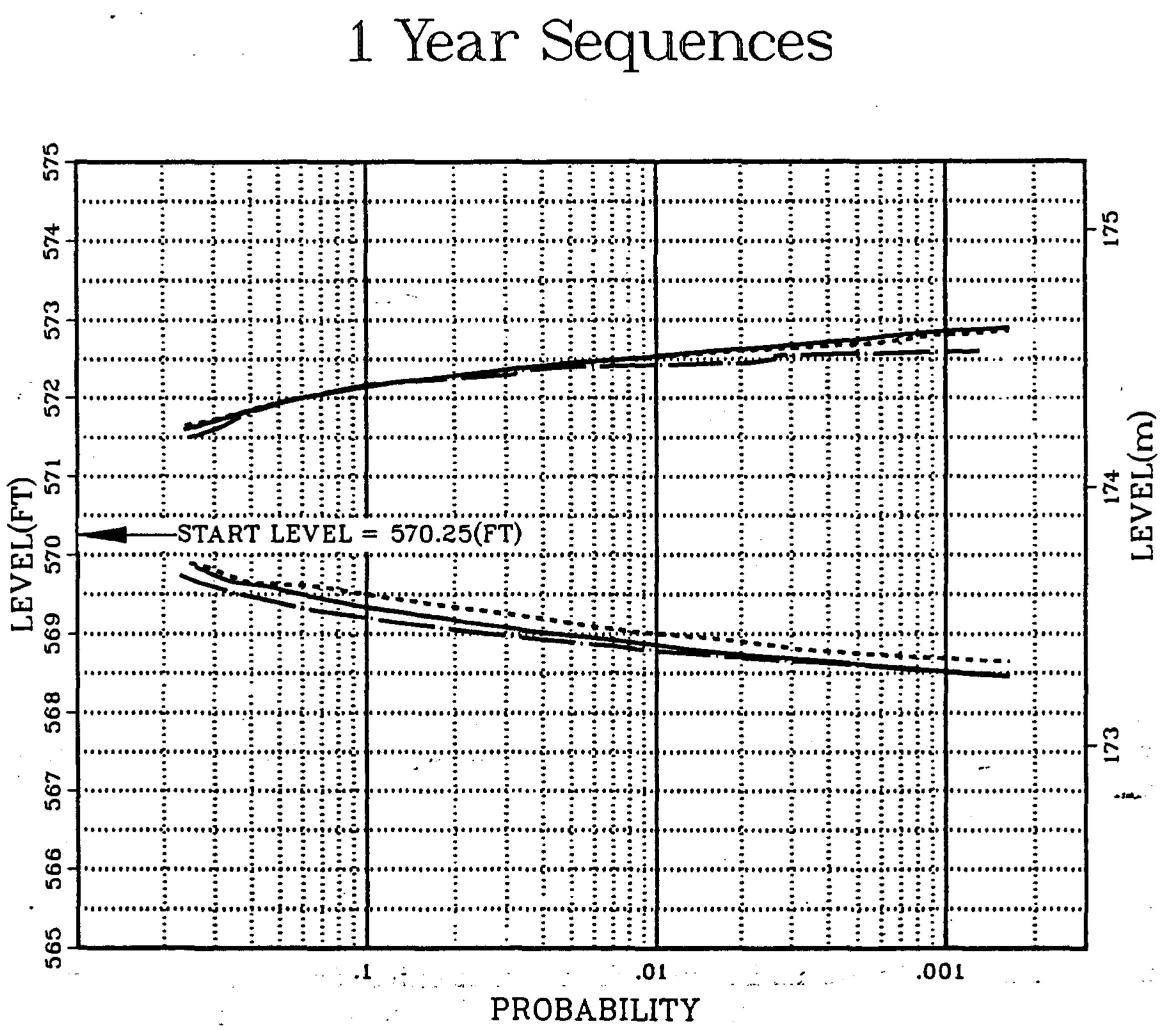

Figure 13. **Example probability distribution for monthly mean Lake Erie levels at Cleveland, Ohio, over a 1-year planning horizon, conditioned on an initial January lake level of 570.25 ft (173.80 m). Solid lines represent no conditioning on climatic regime. Dotted lines represent conditioning on a wet climatic regime, long-dashed lines on a dry regime.**

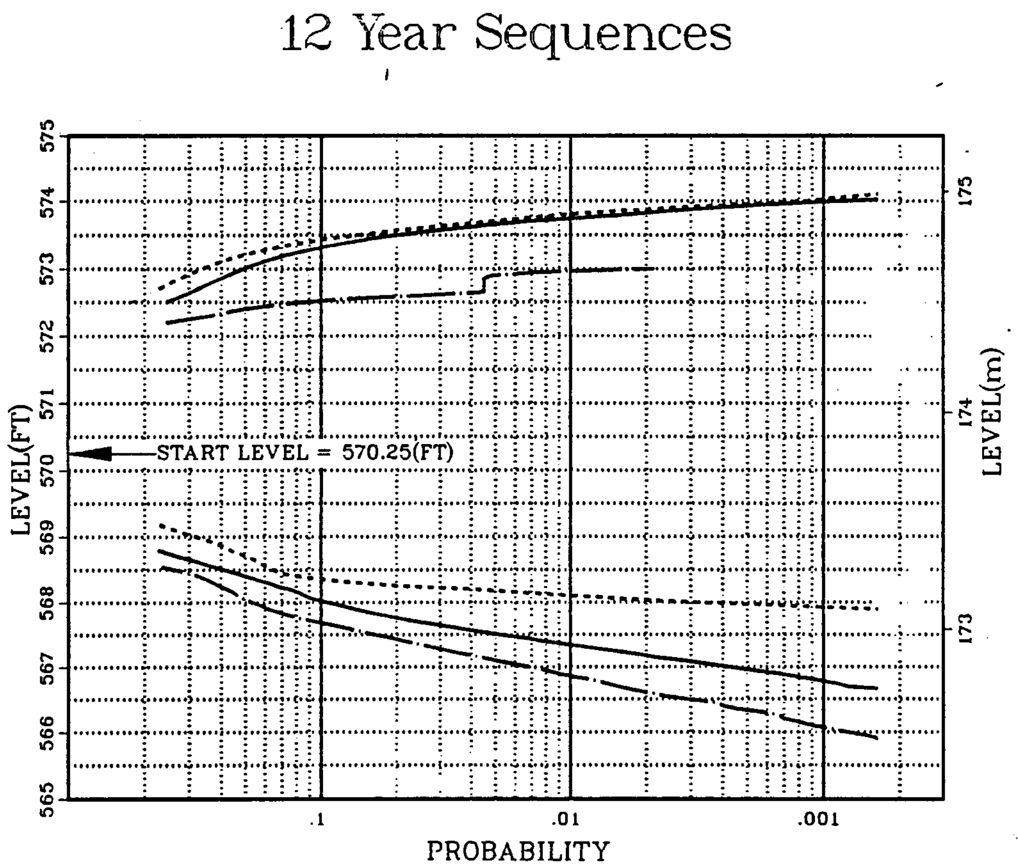

Figure 14. Example probability distribution for monthly mean Lake Erie levels at Cleveland, Ohio, over a 12-year planning horizon, conditioned on an initial January lake level of 570.25 ft (173.80 m). Solid lines represent no conditioning on climatic regime. Dotted lines represent conditioning on a wet climatic regime, long-dashed lines on a dry regime.

Extensions

The preceding analyses show that conditioning by starting level, starting month (seasonality), and wet or dry climatic regime all have important influences on predicted exceedance probabilities, especially in combination with each other. One extension of the methodology which may have important applications is conditioning on various types of lake level trends. It may be useful, for short planning horizons, to condition resampling on the basis of interannual lake level trends over about 5 years, to reflect wetting or drying of basin watersheds resulting from persistent meteorologic conditions within a climatic regime (e.g., due to El Niño effects). For longer planning horizons, conditioning on trends over several decades may permit consideration of climate changes, including shifts among climatic regimes.

Fruitful application of trend conditioning will require objective definitions of the trends, but with sufficient generality that adequate data will be available for resampling from each specified category of conditioning. Objective definitions are required for two reasons. First, analyses made under a particular conditioning require each lake level obtained by resampling to be identified under that definition, in order to determine what category will be subsequently resampled. Second, thoughtful use of the results presumes the ability to identify the condition extant at the start of the planning horizon and estimate the conditions that will prevail during the period. Such definitions for various types of trends might make use of parametric or non-parametric smoothing or measures of coherency such as the Hurst coefficient.

These same caveats pertain to conditioning by levels and seasonality as well. Months and levels are exemplars of objectivity, which is a strong recommendation for their use (as well as the availability of data). It is likely, however, that factors may be found which serve the same function with more predictive power, but for which some objectivity in definition may be sacrificed. Further work on seasonality factors, alternative level-dependent groupings of differences, and alternative definitions of lake level intervals is planned.

While the preceding analyses focused on using monthly mean levels from a centrally located station, our methodology is amenable to extension in several ways. First, it is obviously desirable to "overlay" the effects of shorter-term conditions such as storm- or wind-induced wave activity and seiche effects, particularly for predictions at other locales. This requires using daily or hourly data under parallel conditioning. Analyses of shorter-term effects could then be combined with longer-term monthly mean data results to yield joint conditional probabilities.

Second, data available prior to the historic record used herein may be included, provided they are complete enough to permit the necessary conditioning. On Lake Erie, intermittent monthly mean water level records extend back to 1819 (Tait 1983, Bishop 1987). At a minimum, use of such data requires two consecutive monthly levels and identification of the prevailing climatic regime, in order to add the levels difference to the correct conditional sampling category. This capability greatly extends, in principle, the set of available data since large breaks in data continuity make standard time-series analyses difficult to apply.

A potentially very powerful extension of our methodology is in the range of available statistics. The focus has been on the probability distribution of minima and maxima of lake levels, for given planning horizons, under various conditioning. It is equally straightforward to derive other statistics which might be even more useful to Great Lakes decision makers. Distributions of the time from the start of a planning horizon that some specified lake level is exceeded ("time-to-exceedance" probabilities) would be helpful in developing workable contingency plans, and then implementing elements of those plans as lake level conditions (and thus, probabilities) change. Subsequent analyses using updated levels data could be used to confirm previous analyses or suggest acceleration or postponement of scheduled efforts. Distributions of the length of time a specified level is exceeded during a planning horizon ("duration of exceedance" probabilities) would be useful in developing optimization plans to maximize benefits of lake use even during moderate water level conditions (e.g., for hydropower production). In general, any statistic derivable from a given water level record may be displayed in terms of its distribution, conditioned as before by current starting level, climatic regime, etc. Statistics tailored to the various needs of Great Lakes decision makers could be produced, based on the full extent of historic records and conditioned on the current lake level status.

Conclusions

The implications of our analyses are clear. Development of a single general-purpose probability distribution for monthly mean lake levels is absolutely inappropriate. Probability distributions must consider 1) the length of the planning horizon, 2) the lake level at the start of the planning horizon, 3) the month (or season) of the start of the period, and 4) climatic regimes. For short planning horizons of about 20 years or less, lack of consideration of the first three conditions will result in over-investment in risk reduction measures. As the planning horizon increases beyond 20 years, conditioning on these factors becomes less important, but may still merit consideration when initial levels are extreme and levels of the opposite extreme are of concern. Without explicit consideration of climatic regimes, risks of extreme conditions may be significantly over- or under-estimated; the risk estimates may be especially biased for planning horizons longer than a decade, but the biases may be significant even for shorter periods. Great Lakes levels reflect climatic regimes, and those regimes can have much different implications for potential lake level behavior. Decision makers should consider lake level probabilities for each regime within their decision analysis framework, by weighted combination of probabilities or some other approach. For extremely long planning horizons (e.g., 100 years for massive infrastructures, 1000 years or beyond for lakeshore nuclear plant siting), several climatic shifts might be expected; thus, use of the entire historic record may be sufficient, or preferably, additional conditioning on regime behavior.

A few observations should be made on the fundamentals of the method presented herein. First, resampling statistics do not in themselves provide much perspective into the mechanisms or causative structure of lake levels. No mechanistic modeling is performed to reveal the interrelations of lake levels with precipitation, runoff, evaporation, winds, temperature, human-induced effects, or other processes. Exogenous variables enter the resampling analyses only through the conditioning, based on a priori knowledge of their effects (e.g., grouping levels differences by lake level, on the basis of known lake level-discharge relationships). Research on these variables would be of great benefit in refining the resampling analyses. Second, the data base of approximately 1560 monthly mean levels (and differences), although one of the longest contiguous geophysical records extant for North America, is not really adequate for the resampling analyses imposed on it, even here. As illustrated by Figure 2, even considering the full 130-year historic record, some lake level intervals have very few observations available for resampling; insufficient sample size becomes even more problematic when climatic regimes are considered, as in Figures 11 and 12. It is highly desirable to expand the current historic record with the non-contiguous levels data prior to 1860, to include both periods which are similar to the current record and those which record more extreme conditions. Third, these analyses should be extended to data from other stations and lakes for monthly mean levels as well as daily and hourly levels, to develop conditional lake level probabilities for different locales and time scales.

Resampling analyses modeled on the "bootstrap" method offer an opportunity to use a much larger portion of available data than may be accommodated by existing techniques. The ability to refine probability distributions via conditioning should improve their reliability. The scope of application is increased as well by the ability to tailor results to the specific needs of Great Lakes users. Together these facilities may offer Great Lakes decision makers much greater access to the information contained in historic lake level records and the results of water levels research.

References

Adams, E. E., R. L. Bras, D. R. Harleman, C. Puente, and C. M. Bischoff. 1985. Short term lake level predictions for the Great Salt Lake. In *Problems and Prospects for Predicting Great Salt Lake Levels,* 205-217, P. A. Kay and H. F. Diaz (eds.). Salt Lake City, UT: Center for Public Affairs and Administration, University of Utah.

Bishop, C. T. 1987. *Great Lakes water levels: A review for coastal engineering design.* National Water Research Institute Contribution 87-18. Burlington, ONT: Environment Canada.

Bowles, D. S., and L. D. James. 1985. Issues associated with stochastic modeling of Great Salt Lake levels for planning purposes. In *Problems and Prospects for Predicting Great Salt Lake Levels,* 218-235, P. A.

Kay and H. F. Diaz (eds.). Salt Lake City, UT: Center for Public Affairs and Administration, University of Utah.

Changnon, S. A., Jr. 1987. Climate fluctuations and record-high levels of Lake Michigan. *Bulletin of the American Meteorological Society*. 68:1394-1402.

Efron, B. 1982. *The Jackknife, the Bootstrap, and Other Resampling Plans*. Philadelphia, PA: SIAM.

Federal Emergency Management Agency. 1987. *Questions and Answers on the National Flood Insurance Program*. Washington, D.C.: Federal Insurance Administration.

Guganesharajah, K. and E. M. Shaw. 1984. Forecasting water levels for Lake Chad. *Water Resources Research*. 20:1053-1065.

Hartmann, H. C. 1987. *Potential Variation of Great Lakes Water Levels: A Hydrologic Response Analysis*. NOAA Technical Memorandum ERL GLERL-68. Springfield, VA: National Technical Information Service.

International Joint Commission. 1989. *Living with the Lakes: Challenges and Opportunities, Annex A: Past and Future Water Level Fluctuations*. Washington, D.C.

James, L. D., D. S. Bowles, R. V. Canfield, D. G. Chadwick, Jr., N. Stauffer, and J. P. Riley. 1984. Estimating water surface elevation probabilities for the Great Salt Lake. In *A Critical Assessment of Forecasting in Western Water Resources Management*, 93-108. Bethesda, MD: American Water Resources Association.

Privalsky, V. E. 1981. Fluctuations of water level in a closed lake and their optimal probabilistic prediction. *Vodnye Resursy*. 6:5-21.

Privalsky, V. E. 1988. Modeling long term lake variations by physically based stochastic dynamic models. *Stochastic Hydrology and Hydraulics*. 2:303-315.

Quinn, F. H. 1981. Secular changes in annual and seasonal Great Lakes precipitation, 1854-1979, and their implications for Great Lakes water resource studies. *Water Resources Research*. 17:1619-1624.

Quinn, F. H. 1990. Great Lakes water level statistics and decisionmaking. In *Proceedings of Great Lakes Water Level Forecasting and Statistics Symposium*. Ann Arbor, MI: Great Lakes Commission.

Southeastern Wisconsin Regional Planning Commission. 1989. *A Lake Michigan shoreline erosion management plan for Milwaukee County, Wisconsin*. Community Assistance Planning Report 163. Waukesha, WI: SWRPC.

Tait, B. J. 1983. *Conversion of Historical Lake Erie Water Level Elevations to International Great Lakes Datum (1955)*. Toronto, ONT: Public Works Canada.

U.S. Army Corps of Engineers. *Revised Report on Great Lakes Open-Coast Flood Levels, Phase I*. Detroit, MI: Detroit District.

Wiche, G. J., S. M. Hoetzer, and J. G. Rankl. 1986. *Hydrology of the Devils Lake Basin, Northeastern North Dakota*. Water Resources Investigations Report 3. Bismark, ND: North Dakota State Water Commission.

Woodbury, L. H. and G. Padmanabhan. 1989. Estimating terminal lake level frequencies. *Journal of Water Resources Planning and Management*. 115:321-337.